



A new design and a weighting strategy of a survey on tourist activity of the Croatian population

Lidija Gligorova

Croatian Bureau of Statistics, Zagreb, Croatia
gligoroval@dzs.hr

Marijana Kožul

Croatian Bureau of Statistics, Zagreb, Croatia
kozulm@dzs.hr

Abstract

Until 2014, the sample for the statistical survey Tourist Activity of the Population of the Republic of Croatia, shortly named as TADS, was a quota sample. The sample selection and the data collection for the survey in the Computer Assisted Telephone Interviewing (CATI) mode were conducted by a hired marketing research agency. In the summer of 2014, the Croatian Bureau of Statistics (CBS) set up its own call centre for CATI surveys. At the beginning of 2015, the CBS started developing a new probability sample design and a weighting strategy of TADS. The surveys of all the four quarters of 2015, fully conducted by the CBS, were treated as test surveys. The purpose of this paper is to describe and evaluate the finally accepted overlapping dual-frame survey design and a weighting procedure based on the sample data collected in all the quarters of 2015. The paper shows how the new design is incorporated in the existing circumstances in which surveys are conducted. The overlapping dual-frame survey design of TADS produces survey estimates related to the whole survey population based on two independent samples from two overlapping sampling frames (the first sampling frame consists of landline phone numbers from the phone book and the second one consists of mobile phone numbers from the phone book). In spite of a large undercoverage of the survey population in the phone book, using both phone types in TADS improves the total coverage of the survey population and relieves a problem of coverage and response bias. The applied weighting procedure is described systematically. The probability sampling allows calculating quality indicators such as standard errors of survey estimates and helps optimizing the survey process.

Keywords: composite weight, overlapping dual-frame survey design, probability sample, quota sample, screening dual-frame survey design.

JEL classification: C83.

DOI: 10.1515/crebss-2016-0012

Received: September 02, 2016

Accepted: December 16, 2016

Introduction

Statistical survey Tourist Activity of the Population of the Republic of Croatia, shortly named TADS (the abbrev. TADS comes after the previous survey name in Croatian language: *Turistička aktivnost domaćeg stanovništva*), gathers data on national tourism, that is, the data on participation in tourism and the characteristics of tourism trips of the Croatian citizens aged 15 and over (15+), outside their usual environment (Eurostat, 2015).

The goal of the survey is to assess how many of Croatian citizens are travelling, their motives for going on a trip, where and when they travel, how long do the trips last and what the trip expenditures amount to. Information collected by this survey is related to overnight trips (private and business) and same-day visits (private and business) within Croatia and abroad. While the section on trips includes all tourism trips (private and business), the section on participation in tourism includes only trips for private purposes. Since the business trips are more difficult to recall and because this touches the social component of tourism statistics, the trips made for business reasons are excluded. For the surveyed citizens who did not travel, the data on the reasons for not going on a trip are collected. Except for the previously mentioned characteristics of trips, data are also gathered on the socio-demographic profile of the population (age, gender, level of education, number of persons in the household, status in employment, and county of residence). There is a high national relevance of the survey since it is one of the main data sources for Tourism Satellite Account (TSA) compilation.

The data on tourist activity of the Croatian population in 2015 were gathered by carrying out four waves of Computer Assisted Telephone Interviewing (CATI) survey on a probability sample of Croatian citizens aged 15+. Ideally, the reference period should be one month. However, for the reason of cost-efficiency, a quarter is taken as the reference period. A tourism trip had to end within the reference period and could have started up to 365 days prior to its end. The characteristics of trips are monitored on the level of three overnight stays and three same-day visits as the upper limit.

Merging the results of four quarterly periods gave the assessment of number and characteristics of trips in the observed year. The assessment of the number of persons who went on a trip is calculated based on persons who were surveyed in the last (fourth) survey wave in which the question for the whole reference year were included.

According to Eurostat guidelines, national statistical offices in the European Union member states tend to use probability sampling in their sample surveys. Probability sampling means that every unit in the survey population has a known and greater than zero (inclusion) probability to be selected in the sample. As it is known, by using a probability sampling, a sample statistician avoids subjective judgements (bias) in sample selection. Probability sampling and probability theory ensure proper statistical inference from the sample to the population and possibility to calculate quality indicators such as standard errors.

Probability sample surveys also suffer from coverage and response bias but there are widely accepted different steps to relieve bias problem (Biemer, Lyberg, 2003).

Non-probability sampling is a widely accepted by marketing research agencies (Office for National Statistics, 2015; FAO, 2016). However, national statistical offices sometimes resort to non-probability sampling from different reasons: too high non-response rates and costs, or no ability to conduct probability sampling (Office for National Statistics, 2015).

Until 2014, the sample for TADS was a non-probability quota sample. The quotas were defined taking into account the following variables: county, size of settlement (the measure of size was the number of inhabitants) and gender – age group of respondents. The data collection was conducted by a marketing research agency. The mode of data collection was CATI. The sampled household was selected by random selection of phone number from the landline phone book. The interviewed person at age 15+ in the sampled household was selected using “last – birthday” selection technique, if it was only allowed by the quota approach.

In the summer of 2014, the Croatian Bureau of Statistics (CBS) set up its own call centre for CATI surveys. The first surveys conducted in this centre were agriculture surveys. The phone numbers were selected from the Croatian Agriculture Register.

At the beginning of 2015, the CBS started developing a new (probability) sample design and a weighting strategy of TADS with CATI mode of data collection. All survey quarters in 2015, fully conducted by the CBS, were treated as test surveys because of the complexity of survey questionnaire, survey topics, question difficulties, no skill interviewers and no experience in probability sampling from the phone book.

The purpose of this paper is to describe and evaluate a new overlapping dual-frame survey design and weighting procedure based on sample data collected in all quarters in 2015.

The purpose is also to make potential readers of this paper more familiar with the work in the CBS and encourage better cooperation between the CBS and universities and other institutions. We will show why an overlapping dual sampling frame (landline and mobile phones in the phone book) is better solution as a sampling frame in CATI surveys than a phone book with only landline phones, even in the case of large undercoverage of survey population and phone users in the phone book.

The first idea about sampling frame of TADS was to take a data set of all persons in private households at age 15+ enumerated in the 2011 Census. A person would be a sample unit. The name and address of selected person would be linked to the phone book to take a phone number. However, this approach was not accepted because of many reasons. Both the 2011 Croatian Census data and the phone book are not up-to-date. Information from the phone book is formatted in different ways and it makes linkage difficult. Moreover, in the phone book the coverage of landline phones is much better than the coverage of mobile phones. A landline phone is a way to contact a household and only one name from the household is connected with landline phone number in the phone book (Bethlehem, Cobben, Schouten, 2011). As a result, a large proportion of sampled persons from the 2011 Census would not be linked to the phone book.

On the other hand, a part of phone book would never be linked to the Census data. Since the phone reached population is a target, the phone book is imposed as a sampling frame regardless of the lack of information in it. Variables collected in the survey, and for which distributions in the population are known (e.g., from the Census), can also be used as auxiliary information for weighting or the analysis of survey non-response.

After this introduction, in the following chapter, the assessment of the coverage of survey population and the coverage of all landline and mobile phone users by phone book is given. In addition, differences in socio-demographic characteristics of respondents reached by landline and mobile phones are presented. Further, new sample design of TADS developed in 2015 and accepted weighting scheme of TADS

are described. In addition, finally, some concluding considerations about the influence of applied weighting scheme on final estimates are given.

Research methodology

Sampling frame coverage of survey population

The accessible phone book released by the firm "Imenik" Ltd., contains landline and mobile phone numbers of phone ownerships (users) in Croatia with their names and addresses and information whether the user is defined as private or not.

The target population for TADS is defined as citizens of the Republic of Croatia at age 15+, that live in private households, excluding population living in institutional households, citizens of the Republic of Croatia living abroad and foreign citizens living in Croatia. The aim of this chapter is to assess the coverage of survey population and the coverage of all landline and mobile phone users by phone book, situation existing on last quarter of 2015. In assessment, the 2011 Census data and data released by the Croatian Regulatory Authority for Network Industries (HAKOM) were used.

Tables 2, 3, 4 and Table 5 show number and percentage of sample units according to unit statuses (recorded results of phone calls) after data collection work. Abbr. TADS1501 in the following text means that data (information) are related to the first quarter in 2015 of TADS etc. Status codes used are as given in Table 1. In the Table 3, data for status 12 and status 13 are given together.

Table 1 Status codes

Code	Status	Code	Status
0	Response (selected person is successfully interviewed)	7	Overcoverage - enterprise
1	Refusal	8	Overcoverage – institutional household
2	Refusal to cooperate in any surveys	9	Overcoverage – selected person is younger than 15 (mobile phone)
3	No time	10	Overcoverage - duplicate
4	Illness	11	Other reasons for no cooperation
5	Overcoverage – foreign citizenship	12	No answer (overcoverage?)
6	Overcoverage - craft	13	Overcoverage - Phone number is not used

At the beginning, non-private phone users, users living abroad, duplicate phone numbers, duplicate households with two or more landline phone numbers and duplicate persons with two or more mobile phone numbers were removed from the phone book. After that, there were 1.666.332 phone numbers (users) in the phone book. Thereof there were 1.078.694 landline-phone users, and 587.638 mobile phone users. Among them, about 11% of landline phone numbers and about 30% of mobile phone numbers do not belong to survey population (statuses 5, 6, 7, 8, 9, 10, 13 in Table 2 – Table 5 are given for overcoverage). Considering these percentages, landline phone numbers from the phone books cover about 960,000 private households, and mobile phone numbers cover about 412,000 persons in private households.

HAKOM agency releases data only about total number of landline phone users, not about the number of private landline users.

The number of mobile phone users (post-paid subscribers – they are listed in the phone book) given by HAKOM agency is related to the number of SIM cards. Private mobile phone users according to HAKOM agency requested mobile phone service for private purpose.

Table 2 TADS1501 Sample units according to status codes

Status	0	1	2	3	4	5	6	7	8	9	10	11	12	13	Total
Land-line phone	1,315 41.7	561 17.8	349 11.1	80 2.5	51 1.6	2 0.1	6 0.2	3 0.1	1 0.0	1 0.0	0 0.0	66 2.1	426 13.5	293 9.3	3,154 100%
Mobile phone	335 39.4	108 12.7	49 5.8	31 3.6	11 1.3	1 0.1	0 0.0	0 0.0	0 0.0	2 0.2	0 0.0	25 2.9	33 3.9	256 30.1	851 100%
Total	1,650 41.2	669 16.7	398 9.9	111 2.87	62 1.6	3 0.1	6 0.2	3 0.1	1 0.0	3 0.1	0 0.0	91 2.3	459 11.5	549 13.7	4,005 100%

Source: Author's calculation based on data collected by CBS.

Table 3 TADS1502 Sample units according to status codes

Status	0	1	2	3	4	5	6	7	8	9	10	11	12+13	Total
Land-line phone	1,956 49.3	673 17.0	156 3.9	105 2.7	60 1.5	19 0.5	10 0.3	3 0.1	0 0.0	0 0.0	0 0.0	88 2.2	896 22.6	3,966 100%
Mobile phone	454 43.9	120 11.6	27 2.6	36 3.5	2 0.2	4 0.4	0 0.0	1 0.1	1 0.1	3 0.3	0 0.0	19 1.8	367 35.5	1,034 100%
Total	2,410 48.2	793 15.9	183 3.7	141 2.8	62 1.2	23 0.5	10 0.2	4 0.1	1 0.0	3 0.1	0 0.0	107 2.1	1,263 25.3	5,000 100%

Source: Author's calculation based on data collected by CBS.

Table 4 TADS1503 Sample units according to status codes

Status	0	1	2	3	4	5	6	7	8	9	10	11	12	13	Total
Land-line phone	2,028 51.4	655 16.6	135 3.4	102 2.6	68 1.7	10 0.3	3 0.1	6 0.2	0 0.0	0 0.0	0 0.0	61 1.5	445 11.3	436 11.0	3,949 100%
Mobile phone	489 46.5	119 11.3	18 1.7	30 2.9	9 0.9	3 0.3	0 0.0	0 0.0	0 0.0	1 0.1	0 0.0	13 1.2	31 3.0	338 32.2	1,051 100%
Total	2,517 50.3	774 15.5	153 3.1	132 2.6	77 1.5	13 0.3	3 0.1	6 0.1	0 0.0	1 0.0	0 0.0	74 1.5	476 10.0	764 15.3	5,000 100%

Source: Author's calculation based on data collected by CBS.

Table 5 TADS1504 Sample units according to status codes

Status	0	1	2	3	4	5	6	7	8	9	10	11	12	13	Total
Land-line phone	2,351 51.9	591 13.0	206 4.6	82 1.8	64 1.4	3 0.1	21 0.5	16 0.4	7 0.2	0 0.0	7 0.2	67 1.5	525 11.6	592 13.1	4,532 100%
Mobile phone	1,319 53.4	194 7.9	71 2.9	57 2.3	14 0.6	2 0.1	1 0.0	2 0.1	2 0.1	11 0.5	1 0.0	72 2.9	47 2.0	677 27.4	2,470 100%
Total	3,670 52.4	785 11.2	277 4.0	139 2.0	78 1.1	5 0.1	22 0.3	18 0.3	9 0.1	11 0.2	8 0.1	139 2.0	572 8.2	1,269 18.1	7,002 100%

Source: Author's calculation based on data collected by CBS.

Table 6 shows coverage of survey population regarding landline phones, and Table 7 shows coverage of survey population regarding mobile phones.

Table 6 Landline phone – coverage of survey population

Estimated No. of private users (households) - phone book (A)	No. of users- HAKOM (B)	A/B (%)	No. of private households – the 2011 Census (C)	A/C (%)
960,000	1,314,654	73.0	1,519,038	63.2

Source: Author's calculation based on data collected by CBS, HAKOM.

Table 7 Mobile phone – coverage of survey population

Estimated No. of private users (persons) – phone book (A)	No. of private users – post-paid subscribers – HAKOM (B)	A/B (%)	No. of persons at age 15+ in private households – the 2011 Census
412,000	1,300,763	31.7	3,632,461

Source: Author's calculation based on data collected by CBS, HAKOM.

Since the number of landline phone users released by HAKOM agency is related to all users (not only private), we can assume, from Table 6, a quite good coverage of private landline phone users by phone book (over 73.0%) and over 63% of coverage of survey population (private households).

Mobile phone numbers in the phone book improve the survey population coverage.

In spite of the undercoverage of landline and mobile phone users in the phone book, we can assume that the presence of any valid phone number in the phone book is 'random enough' to prevent significant coverage bias (probably it is not true, but coverage bias will be topic for new research in the future).

Maybe it was expected that the survey data collection in all quarters of 2015 would indicate significant differences between landline and mobile phones in the sample, which are presented further in this chapter.

The interviewed person at age 15+ in the sampled household (landline phone) was selected using "last – birthday" selection method.

Response rate (RR) in TADS is defined as:

$$RR = \frac{\text{No. of valid units successfully interviewed}}{\text{No. of valid contacted units}} = \frac{\sum [Status = 0]}{\sum [Status = 0 + 1 + 2 + 3 + 4 + 11]} \quad (1)$$

Table 8 Response rate and percentage of units, which have a trip

Reference period	Phone type	Response rate (%)	No. of responding units (status=0)	Responding units which have a trip	
				No.	%
TADS1501	Total	55.4	1,650	381	23.1
	Landline phone	54.3	1,315	258	19.6
	Mobile phone	60.0	335	123	36.7
TADS1502	Total	65.2	2,410	815	33.8
	Landline phone	64.4	1,956	636	32.5
	Mobile phone	69.0	454	179	39.4
TADS1503	Total	67.5	2,517	1,289	51.2
	Landline phone	66.5	2,028	1,007	49.7
	Mobile phone	72.1	489	282	57.7
TADS1504	Total	72.1	3,670	1,421	38.7
	Landline phone	69.9	2,351	804	34.2
	Mobile phone	76.4	1,319	617	46.8

Source: Author's calculation based on collected data by CBS.

Table 8 shows a higher response rate from the quarter to quarter of survey as a result of improved data collection work. The response rate is higher for mobile phones than landline phones. In addition, the percentage of interviewed persons who have a trip in the reference period is higher for mobile phones.

Statuses 12 and 13 are not treated as valid units in formula (1). High proportion of status 12 for landline phones (Table 2 – Table 5) could be explained with a decreasing trend of using landline phone services and a continued growth trend of using mobile phone services (Ekonomski institut Zagreb, 2015).

High proportion of status 13, particularly for mobile phones, indicates that the phone book is not up-to-date.

Table 9 shows distributions of interviewed persons (status=0) according to phone type and different socio-demographic characteristics for all quarters in 2015. This Table again confirms that there are differences between respondents reached by landline and mobile phone.

Table 9 Socio-demographic characteristics of respondents

	TADS1501		TADS1502		TADS1503		TADS1504	
Socio-demograph. category	Landl. phone (%)	Mobile phone (%)	Landl. phone (%)	Mobile phone (%)	Landl. phone (%)	Mobile phone (%)	Landl. phone (%)	Mobile phone (%)
By gender								
Male	38.2	50.1	40.4	51.5	35.1	51.1	36.3	52.5
Female	61.8	49.9	59.6	48.5	64.9	48.9	63.7	47.5
By age groups								
15-29	10.8	15.5	12.3	13.0	9.5	17.4	8.3	21.3
30-49	23.3	40.3	23.7	40.5	23.7	38.2	28.7	38.4
50-64	27.6	31.1	28.6	33.0	31.3	30.7	28.9	27.3
65+	38.3	13.1	35.4	13.5	35.5	13.7	34.1	13.0
By highest completed level of education								
Primary school or lower	25.1	8.7	24.2	8.1	24.7	7.9	17.4	7.3
Secondary school	50.8	59.1	52.4	60.4	52.4	61.8	55.5	58.8
College	12.6	15.8	11.0	13.2	11.6	13.9	12.3	14.6
Faculty or higher	11.5	16.4	12.4	18.3	11.3	16.4	14.8	19.3
By status in employment								
Employed	28.2	54.3	35.2	54.8	30.4	55.6	34.6	53.1
Unemployed	14.7	15.2	10.8	15.0	15.4	16.8	14.2	16.7
Student	5.5	4.8	6.2	4.6	4.1	4.3	3.8	6.8
Others	51.6	25.7	47.8	25.6	50.1	23.3	47.4	23.4
By household size groups								
1	19.5	12.5	21.5	15.0	21.1	11.5	20.3	14.0
2	30.2	23.3	28.0	22.5	29.7	22.3	28.8	24.0
3	17.2	22.4	17.8	22.9	17.2	23.7	17.9	22.8
4	19.1	22.1	18.4	24.0	18.0	25.6	19.6	23.9
5 or more	14.0	19.7	14.3	15.6	14.0	16.9	13.4	15.3

Source: Author's calculation based on collected data by CBS.

Table 10 again shows distributions of interviewed persons according to gender and age groups for landline and mobile phones for the last quarter in 2015. Added distributions in the last column are taken from the 2011 Census and related to survey population (persons at the age 15+ in private households).

Table 10 Socio-demographic characteristics of respondents – Comparison with 2011 Census data

Socio-demographic characteristics	TADS1504		2011 Census
	Landline phone (%)	Mobile phone (%)	Survey population (%)
By gender			
Male	36.3	52.5	47.7
Female	63.7	47.5	52.3
By age groups			
15-29	8.3	21.3	21.9
30-49	28.7	38.4	32.3
50-64	28.9	27.3	24.9
65+	34.1	13.0	20.9

Source: Author's calculation based on collected data by CBS, Census 2011.

Distributions for landline and mobile phones complement each other. Using both in the sample reduces undercoverage and non-response errors.

In the last quarter of 2015, interviewed persons reached by landline phone were asked whether they had a mobile phone and interviewed persons reached by

mobile phones were asked whether they had a landline phone. Collected information is given in Table 11.

Table 11 Collected data on question about another phone type

Question	TADS1504		
	Answer	Landline phone (%)	Mobile phone (%)
Do you have a mobile phone?	Yes	77.9	-
	No	22.1	-
Do you have a landline phone?	Yes	-	79.0
	No	-	21.0

Source: Author's calculation based on data collected by CBS.

According to Table 11, there is a similar proportion of users of both phone types (dual-phone users) among interviewed persons reached by landline and reached by mobile phone, but, according to previous tables, they are not the same considering their socio-demographic characteristics. It is established that there are two groups of dual-phone users: one group is prone to using landline phone and another group is prone to using mobile phone. It was the main reason why the overlapping dual-frame sample design described in the next chapter was finally used in TADS.

In spite of large undercoverage of landline phone and mobile phone users and large undercoverage of survey population in phone book, using both phone types in TADS improves total coverage of survey population and relieve problem of coverage and response bias.

Sample design

For the first three quarters in 2015, survey design of TADS was screening dual-frame design (Brick et al., 2011). For the last quarter in 2015, sample design was changed to overlapping dual-frame design. The term dual-frame means that there are two frames: one with landline phone numbers and another with mobile phone numbers. Usually such designs are connected with the Random Digit Dialling (RDD) technique (Brick, 2008; HAKOM, 2015) of selecting persons by generating phone numbers at random. We applied dual-frame design with the phone book as sampling frame (two (sub) frames).

With the screening dual-frame design, the aim was to have the frame coverage of survey population as good as possible and without overlapping. Non-private users, duplicate households with two or more landline phone numbers, duplicate persons with two or more mobile phone numbers and persons with mobile phone number also covered by landline phone number according to the their names and addresses, were removed from the phone book (screening). The frame covered landline phone users who have landline phone number listed in the phone book and mobile phone users who have only mobile phone number listed in the phone book.

With the overlapping dual-frame design, only non-private users and duplicate phone numbers were eliminated from the frame. In this case, there are two (sub) frames (landline and mobile phone frames) which cover landline phone users who have only landline phone number listed in the phone book (landline phone frame), mobile phone users who have only mobile phone number listed in the phone book (mobile phone frame), and dual-phone users (overlapping part of landline and mobile phone frames), who have landline and mobile phone number listed in the phone book. The person with a mobile phone number in the phone book and also covered by landline phone number in phone book could be selected by selection

either her/his the landline or mobile phone number which makes weighting procedure more complex.

The purpose of overlapping dual-frame survey design is to produce survey estimates related to the whole survey population based on two independent samples from two overlapping sampling frames.

In all quarters in 2015, the sample was a stratified random sample with stratification of sampling frame according to phone type and county where phone user lives. Sample allocation was proportional to the number of units in stratum. In the first quarter, sample size was 4,005 units. Because of low response rate, in the second and third quarters, sample size was increased to 5,000 units. In the last quarter, sample size was 7,002 units.

There is no overlapping of sample units (phone numbers) between different quarters of survey. A person who is multiple respondent in the survey (e.g. a person is selected via landline phone and mobile phone, a person has more than one mobile phone) is interviewed only once. If such person is contacted again in the same wave of survey, he/she gets status code 10 (duplicate unit) for the second approach.

In the case of landline phone in the sample, the interviewed person at age 15+ in the sampled household was selected using "last – birthday" selection method.

In the case of mobile phone in the sample, the interviewed person is phone user.

Calculation of weights

In sample surveys, every responding sample unit represents a certain number of non-sampled units given by weight. The main reasons for using weights in sample surveys, theoretical background and practical applications in the case of dual-frame survey design are given e.g. by Barr et al. (2014), Elkasabi (2014), Foreman (1991), Gabler et al. (2012), Kish (1965). Here, we will mention that sample units / sample data are weighted because of different inclusion probability of units and non-response, overcoverage and undercoverage bias. Using appropriate weighting, we produce more accurate survey estimates of population parameters.

In this chapter, calculation of sample weights applied in TADS in the last quarter in 2015 is described.

Design weights w_{1h} (reverse of probability that a given sample includes a specified unit) joined to every selected unit in the sample in stratum h , $h=1,..., H$, are calculated as:

$$w_{1h} = \frac{N_h}{n_h}, \quad (2)$$

where N_h is the number of units in the sampling frame in stratum h , n_h is the number of units in the selected sample in stratum h .

Other weights given further are calculated only for responding units (status=0). For other units they have a value 0.

Non-response weights w_{2h} are calculated with a stratum as a weighting class because of lack of information from sampling frames for more sophisticated non-response correction:

$$w_{2h} = \frac{\sum [Status=0+1+2+3+4+11+12]_h}{\sum [Status=0]_h}. \quad (3)$$

The poststratification weights, w_{3Mh} which relieve the problem of mobile phone undercoverage, are calculated only for units from mobile phone frame:

$$w_{3Mh} = \frac{N_{MPOST}}{\sum_{i=1}^{H_M} (\sum [Status=0]_i) \times w_{1i} \times w_{2i}}, \quad (4)$$

where N_{MPOST} is the number of private mobile phone users (post-paid subscribers) according to HAKOM Agency H_M is related to the number of strata in mobile phone frame.

The postratification weights, which relieve the problem of landline phone undercoverage, are not calculated because of the lack of information about the number of private landline phone users in population.

In selected household i (landline phone), interviewed person is selected with probability $1/p_i$ where p_i is the number of persons in household i at age 15+. Hence, the next weight w_{4Lhi} is defined for units in stratum h in landline phone frame as

$$w_{4Lhi} = p_i. \quad (5)$$

For interviewed person i in stratum h who has only landline phone number or only mobile phone number listed in the landline or mobile phone frame, final weights before calibration w_{Lhi} (landline phones) and w_{Mhi} (mobile phones) are defined as

$$w_{Lhi} = w_{1h} \times w_{2h} \times w_{4Lhi}, \quad (6)$$

$$w_{Mhi} = w_{1h} \times w_{2h} \times w_{3Mh}. \quad (7)$$

A certain proportion of interviewed persons have both landline phone number and mobile phone number listed in frames (because of overlapping part of sampling frames). For those persons, the sample weights w_{Lhi} or w_{Mhi} given also with formulas (6) and (7) were adjusted using the composite factor λ set at 0.5 (Barr et al., 2014; Brick et al., 2011). Therefore, for those dual phone-users, the composite weights (final weights before calibration) are given by the following formulas:

$$w_{Lhi}^\lambda = \lambda \times w_{Lhi}, \quad (8)$$

for interviewed persons selected from landline phone frame and

$$w_{Mhi}^\lambda = (1 - \lambda) \times w_{Mhi}, \quad (9)$$

for interviewed persons selected from mobile phone frame.

Dual-phone users described above were identified with the following questions:

- a) Do you have landline/mobile phone number?
- b) Is this number listed in the phone book?

A certain proportion of respondents do not know whether her/his phone number is listed in the phone book and then this information is imputed randomly.

By the calibration procedure, the weights are adjusted so that they reproduce the totals of external variables. For the calibration of weights given with formulas (6), (7), (8) and (9), we used SAS Macro Calmar (truncated linear method; INSEE, 2002).

The new weights were calibrated on the number of persons in private households at age 15+ by gender, age groups and counties and according to the highest completed level of education. The weights of one-person households were calibrated on the number of persons in one-person households by age groups. External data was taken from the 2011 Census data.

Sample errors for the main estimates were calculated applying the Taylor linearization technique (SAS SURVEYMEANS procedure, complex design).

Results, discussion and considerations for further research

Postratification weights w_{3Mh} for mobile phones are given with formula (4). In the same way, we can define postratification weights w_{3Lh} for landline phones, as follows:

$$w_{3Lh} = \frac{N_{LPOST}}{\sum_{i=1}^{H_L} (\sum [Status=0]_i) \times w_{1i} \times w_{2i}}, \quad (10)$$

where N_{LPOST} is the number of landline phone users according to an external source, H_L is related to the number of strata in landline phone frame.

Here, we want to show to what extent final survey estimates (estimates after calibration of weights) depend on values of N_{LPOST} and N_{MPOST} in the case of TADS on the data from the last quarter in 2015. The weighting procedure is described in the previous chapter. We will change only the formula (6):

$$w_{Lhi} = w_{1h} \times w_{2h} \times w_{3Lh} \times w_{4Lhi}. \quad (11)$$

In the previous chapter, the weight $w_{3Lh} = 1$.

In our consideration, we will distinguish the following interesting cases, as given in Table 12, with comments.

Table 12 Considered cases

Case	Weights properties	Comment
A	$w_{3Lh} = 1, w_{3Mh} = 1$	postratification weights are not applied
B	$w_{3Lh} = 1, N_{MPOST} = 1,300,763$	HAKOM value for N_{MPOST}
C	$w_{3Lh} = 1, N_{MPOST} = 800,000$	arbitrary value for N_{MPOST}
D	$N_{LPOST} = 1,300,000, N_{MPOST} = 1,300,763$	arbitrary value for N_{LPOST} , HAKOM value for N_{MPOST}
E	$w_{3Lh} = 1, w_{3Mh} = 0$	mobile phones are excluded

Source: Author's creation.

Comparison of survey estimates, TADS1504, are shown in Table 13.

Without elaborating the algorithm for truncated linear method for calibration, Table 13 shows that postratification weights (w_{3Lh} and w_{3Mh}), or the extent of sampling frame undercoverage, described above, have a small influence on final point estimates which are close to each other (cases A, B, C, D) but influence the coefficient of variation of calibrated weights and sample errors (the width of 95% confidence intervals). It indicates that different sampling fractions for different frames in dual-frame survey design should be taken into account for optimal sample allocation.

Case E, Table 12, is related only to landline phone sample.

Table 13 Comparison of survey estimates, TADS1504

	Cases				
	A	B	C	D	E
n – sample size (Status=0)	2,912	2,912	2,912	2,912	1,916
CV for calibrated weights (%)	117.0	93.1	101.9	99.3	114.7
Number of persons on trips					
Estimated value	1,108,177	1,106,188	1,114,314	1,113,584	1,134,583
95% confidence interval - Lower limit	1,000,688	1,015,111	1,015,361	1,016,222	992,399
95% confidence interval - Upper limit	1,215,667	1,197,265	1,213,268	1,210,946	1,276,767
Number of persons on trips with overnight stays					
Estimated value	566,162	564,850	566,316	566,251	579,077
95% confidence interval - Lower limit	484,684	496,454	492,524	494,166	476,703
95% confidence interval - Upper limit	647,640	633,246	640,109	638,335	681,451
Number of persons on private trips with overnight stays					
Estimated value	511,432	511,930	513,195	514,502	526,583
95% confidence interval - Lower limit	432,230	445,459	441,345	444,244	426,760
95% confidence interval - Upper limit	590,635	578,401	585,045	584,759	626,406
Number of persons on business trips with overnight stays					
Estimated value	72,472	73,615	72,591	71,529	66,968
95% confidence interval - Lower limit	48,244	52,109	50,652	50,155	38,419
95% confidence interval - Upper limit	96,670	95,122	94,530	92,904	95,516
Number of same-day visitors					
Estimated value	643,517	632,095	642,980	638,932	652,087
95% confidence interval - Lower limit	563,744	563,236	568,626	565,619	545,802
95% confidence interval - Upper limit	723,290	700,953	717,333	712,246	758,373
Number of same-day visitors on private trips					
Estimated value	597,862	587,509	597,393	594,639	610,874
95% confidence interval - Lower limit	520,134	520,293	524,816	522,921	506,482
95% confidence interval - Upper limit	675,590	654,726	669,971	666,356	715,267
Number of same-day visitors on business trips					
Estimated value	82,461	74,080	78,831	74,904	77,734
95% confidence interval - Lower limit	54,928	51,619	53,921	51,945	48,120
95% confidence interval - Upper limit	109,994	96,542	102,842	97,862	107,348

Source: Author's calculation based on collected data by CBS.

In case B, the coefficient of variation of calibrated weights is the smallest. Weighting procedure applied in TADS, corresponds with the case B.

Conclusions

Since 2015, TADS survey has been fully conducted by the CBS with a new overlapping dual-frame survey design and a new weighting strategy based on probability sample of landline and mobile phone users in contrast to quota sample of landline phone users in previous years. Lacking better sampling frame for CATI mode of data collection, the sampling frame is a phone book with all its limitations, divided in two parts (two (sub) frames): landline and mobile phone numbers.

The aim of overlapping dual-frame survey design is to produce survey estimates related to the whole survey population based on two independent samples from two overlapping sampling frames.

This paper gives the assessment of the coverage of survey population and the coverage of all landline and mobile phone users by phone book.

The response rate is higher for mobile phones than landline phones.

Moreover, the survey data collection in all quarters of 2015 indicates significant differences between landline and mobile phones in the sample.

There are two groups of dual-phone users (users of both phone types): one group is prone to using landline phone and another group is prone to using mobile phone.

In spite of large undercoverage of landline phone and particularly mobile phone users and large undercoverage of survey population in phone book, using both phone types in TADS improves total coverage of survey population and relieves the problem of coverage and response bias.

The paper describes the applied weighting procedure step-by-step, which takes into account the main principle of weighting: a weight is a reverse of probability of unit selection. The person with a mobile phone number in the phone book and covered by landline phone number in phone book has multiple chances to be selected. This problem is relieved using the composite weights.

Finally, weights reproduce the totals of selected external variables from the 2011 Census (calibration).

The paper gives the table with some survey estimates and related 95% confidence intervals to show that estimates depend on the extent of sampling frame undercoverage / sampling fractions.

The overlapping dual-frame survey design is used in surveys with CATI mode of data collection in national statistical offices in many countries. This paper describes how this survey design is applied in the CBS and gives opportunity to potential readers to evaluate our work.

References

1. Barr, M. L., Ferguson, R. A., Hughes, P. J., Steel D. G. (2014). Developing a weighting strategy to include mobile phone numbers into an ongoing population health survey using an overlapping dual-frame design with limited benchmark information. *BMC Medical Research Methodology*, Vol. 14, No. 1, pp. 102-112.
2. Bethlehem, J., Cobben, F., Schouten, B. (2011). *Handbook of nonresponse in household surveys*. Wiley, Hoboken.
3. Biemer, P. P., Lyberg, L. E. (2003). *Introduction to Survey Quality*. John Wiley & Sons, Hoboken.
4. Brick, J. M., Cervantes, I. F., Lee, S., Norman, G. (2011). Nonsampling errors in dual frame telephone surveys. *Survey Methodology*, Vol. 37, No. 1, pp. 1-12.
5. Brick, J. M. (2008). *Random Digit Dialing*. In *Encyclopedia of Survey Research Methods*, P. J. Lavrakas (ed.), SAGE Publications Inc. Available at <http://sk.sagepub.com/reference/survey/n436.xml> [08 April 2015].
6. Ekonomski institut Zagreb (2015). *SA Sektorske analize*. Available at <http://www.eizg.hr/hr-HR/Sektorske-analize-o-telekomunikacijama-1477.aspx> [02 July 2015].
7. Elkasabi, M. A. (2014). *Weighting Procedures for Dual Frame Telephone Surveys: A Case Study in Egypt*. Available at http://surveyinsights.org/wp-content/uploads/2014/11/Weighting-procedures-for-dual-frame-telephone-surveys_PAPER-Egypt-Case-study-RESUBMIT1.pdf [14 December 2014].
8. Eurostat (2015). *Methodological manual for tourism statistics – Version 3.1 – 2014 edition*. Available at <http://ec.europa.eu/eurostat/web/products-manuals-and-guidelines/-/KS-GQ-14-013> [07 January 2015].
9. FAO (2016). *Chapter 7: Sampling In Marketing Research*. Available at <http://www.fao.org/docrep/W3241E/w3241e08.htm> [11 March 2016].
10. Foreman, E. K. (1991). *Survey Sampling Principles*. Dekker, New York.
11. Gabler, S., Häder, S., Lehnhoff, I., Mardian, E. (2012). Weighting for Unequal Inclusion Probabilities and Nonresponse in Dual Frame Telephone Surveys. In *Telephone Surveys in Europe*, Häder, S., Häder, M., Kühne, M. (Eds.), Springer – Verlag, Heidelberg, pp. 147-167.
12. HAKOM (2016). *Tromjesečni usporedni podaci tržišta elektroničkih komunikacija u Republici Hrvatskoj Četvrto tromjesečje 2015. godine*. Available at https://www.hakom.hr/UserDocImages/2016/e_trziste/Tromjesečni%20usporedni%20podaci%20za%20tržište%20elektroničkih%20komunikacija%20RH,Q42015.pdf [02 March 2016].

-
13. INSEE (2002). *La macro CALMAR*. Available at <http://www.insee.fr/fr/methodes/outils/calmar/doccalmar.pdf> [19 March 2002].
14. Kish, L. (1965). *Survey Sampling*. John Wiley & Sons, New York.
15. Office for National Statistics (2015). *Non-probability Survey Sampling in Official Statistics*. Available at <https://www.ons.gov.uk> [19 June 2015].
16. HAKOM (2015). *Pojmovnik*. Available at https://www.hakom.hr/UserDocsImages/2015/e_trziste/IZ-AT-UT-OPR-Pojmovnik-v10.pdf [07 April 2015].
-

About the authors

Lidija Gligorova, B.Sc. in Mathematics, graduated at Faculty of Science, Department of Mathematics, University of Zagreb. She works as a senior adviser – specialist in the Croatian Bureau of Statistics, her interest areas in statistics are sampling, unit and item non-response with imputation of missing values and survey quality. The author can be contacted at gligoroval@dzs.hr

Marijana Kožul, B.Sc. in Economics, graduated at Faculty of Economics and Business, University of Zagreb. She works as a senior adviser in the Croatian Bureau of Statistics, her interest areas in statistics are survey methodology and quality assurance/improvement of statistical production processes in tourism sector. The author can be contacted at kozulm@dzs.hr